

# Greedy Sampling and Incremental Surrogate Model-based Tailoring of Aeroservoelastic Model Database for Flexible Aircraft

Yi Wang<sup>1</sup>

*University of South Carolina, Columbia, SC 29201*

Kapil Pant<sup>2</sup>

*CFD Research Corporation, Huntsville, AL 35806*

*and*

Martin J. Brenner<sup>3</sup>, Jeffrey A. Ouellette<sup>4</sup>

*NASA Armstrong Flight Research Center, Edwards, CA 93523*

This paper presents a data analysis and modeling framework to tailor and develop linear parameter-varying (LPV) aeroservoelastic (ASE) model database for flexible aircrafts in broad 2D flight parameter space. The Kriging surrogate model is constructed using ASE models at a fraction of grid points within the original model database, and then the ASE model at any flight condition can be obtained simply through surrogate model interpolation. The greedy sampling algorithm is developed to select the next sample point that carries the worst relative error between the surrogate model prediction and the benchmark model in the frequency domain among all input-output channels. The process is iterated to incrementally improve surrogate model accuracy till a pre-determined tolerance or iteration budget is met. The methodology is applied to the ASE model database of a flexible aircraft currently being tested at NASA/AFRC for flutter suppression and gust load alleviation. Our studies indicate that the proposed method can reduce the number of models in the original database by 67%. Even so the ASE models obtained through Kriging interpolation match the model in the original database constructed directly from the physics-based tool with the worst relative error far below 1%. The interpolated ASE model exhibits continuously-varying gains along a set of prescribed flight conditions. More importantly, the selected grid points are distributed non-uniformly in the parameter space, a) capturing the distinctly different dynamic behavior and its dependence on flight parameters, and b) reiterating the need and utility for adaptive space sampling techniques for ASE model database compaction. The present framework is directly extendible to high-dimensional flight parameter space, and can be used to guide the ASE model development, model order reduction, robust control synthesis and novel vehicle design of flexible aircraft.

## Nomenclature

$A$	= state matrix
$B$	= input matrix
$C$	= output state matrix
$D$	= input transition
$F$	= the universal Kriging model
$G$	= dynamic system
$H$	= design matrix for regression
$N_d$	= dimension of flight parameter space
$q$	= pitch rate

---

<sup>1</sup> Associate Professor, University of South Carolina, AIAA Member; [yiwang@cec.sc.com](mailto:yiwang@cec.sc.com)

<sup>2</sup> Vice President, CFD Research Corporation, non AIAA Member

<sup>3</sup> Aerospace Engineer, Aerostructures Branch, and AIAA Senior Member

<sup>4</sup> Aerospace Engineer, Aerostructures Branch, and AIAA Member

$\mathbf{R}_{krig}$	= Spatial correlation function matrix
$R_{krig}$	= Entries of spatial correlation function matrix
$\mathbf{S}$	= Sampled grid point bin
$\mathbf{U}$	= Unsampled grid point bin
$t$	= time
$d_n$	= spatial distance
$e$	= relative error
$\mathbf{f}$	= ordinary Kriging model to capture local variation
$\bar{\mathbf{f}}$	= global trend function
$\mathbf{g}$	= basis functions/regressors of the low-order polynomial regression for the trend function
$\mathbf{r}_{krig}$	= vector storing spatial correlation between selected and unselected grid points
$s$	= selected sample
$u$	= input signals
$x$	= state
$y$	= output
$\mathbf{w}$	= regression coefficients
$\varepsilon$	= tolerance of the relative error
$\theta$	= correlation coefficient
$\rho$	= a vector of measurable parameters
$\omega$	= angular frequency
$\Re$	= spatial correlation function in a single dimension
$\sim$	= quantities at the sampled/selected grid points
$\wedge$	= quantities at the unsampled/unselected grid points

## I. Introduction

The flight performance of aerospace systems is governed by the interaction between aerodynamics, structural dynamics, and flight control dynamics. Modern flexible aircrafts utilize state-of-the-art materials and light weight and long span designs to achieve better maneuverability and higher performance, e.g., fuel efficiency and range. The lighter and more flexible structures give rise to significant complexity arising from flight dynamics, stability, and durability issues. Instabilities, such as flutter, limit cycle oscillations (LCO) [1], and gust loads [2] as a result of the interplay among control systems, aerodynamic, and structural nonlinearities can lead to unacceptable flight conditions and risk to the mission. Therefore, the ability to accurately model, simulate, and predict aeroelastic (AE) behavior in conjunction with the flight control system (sensors, actuators, and control law) is essential for developing lightweight and high-performance vehicles. In contrast to the rigid aircraft, ASE analysis of the flexible aircraft requires high-fidelity and multiphysics models involving coupled aerodynamics, rigid-body and structure dynamics, and sensor/actuator dynamics. This becomes even more challenging when effects of multi-dimension flight parameters such as air speed, fuel weight, etc. need to be taken into account.

One of the most widely used approaches to analyze aeroservoelastic (ASE) of flexible aircraft is to construct a large number of linear time invariant (LTI) models with refined grid points to accommodate the entire flight envelope with the model matrices on the grid points dependent on flight parameters. While practically feasible given the present computing and software technologies and platforms, construction and use of a large LTI model database can give rise to critical issues that may potentially overwhelm the improvement in modeling accuracy, such as significant burdens of model order reduction (MOR) [3-11] and flight control design [3, 12, 13] as the controller must be synthesized for LTI model at each grid point. In particular, for switched linear parameter varying (LPV) controls, a small number of LTI models are preferred in order to reduce the number of controller switching, and hence, mitigate high order dependency on scheduling parameters [14]. In addition, when common projection/transformation is used to achieve consistent state representation in MOR of LPV ASE models [5, 8, 15], distribution of grid points can also alter the matrix to be projected on and impact ultimate performance of reduced order models. From the implementation perspective, given the practical limit of the size of onboard computer memory, judicious use of a smaller set of models that well resolve regions of rapid-varying flight dynamics is more desirable and cost effective for LPV flight control of flexible aircraft. Therefore a mathematically rigorous approach that can be used to either guide compact ASE model database development or to tailor any existing database to fulfill the same goal in an automated manner is needed. The significance has been clearly identified in a recent study by Al-Jiboory *et al.* [14] for LPV model development for a flexible wing aircraft, in which a method of adaptive grid

step size determination for the database tailoring was developed and demonstrated using the NASA Generic Transport Model within 1D parameter space (Mach).

This paper presents a data analysis framework based on automated sampling of important grid points and incremental surrogate modeling to tailor and develop LPV ASE model database for flexible aircrafts in broad 2D flight parameter space. The ASE model database of the aircraft with flexible wings currently being tested at NASA/AFRC for flutter suppression and gust load alleviation is used as an example for analysis, verification, and demonstration. Specifically, the Kriging surrogate model is developed using ASE models at a fraction of grid points within the original database of the aircraft, which then can be used to obtain models at arbitrary flight conditions through surrogate model interpolation. The greedy algorithm is exploited to select the next sample point that carries the worst relative error between the surrogate model prediction and the benchmark model in the frequency domain among all input-output channels. The process is iterated to incrementally refine and update the surrogate model till a pre-determined prediction tolerance or iteration budget is met. The present method not only automates the grid point selection by eliminating the empirical, trial-and-error process for database compaction, but also identify regions within flight parameter space featuring distinctly different dynamic behavior that should be given adequate consideration during modeling, analysis, and design stages. Different from previous approaches of adaptive grid step size determination (AGSSD) [7] that essentially relies on calculation of  $H_{(\sigma_2)}$ -norm of the model error along a predetermined direction in the flight parameter, our method utilizes the greedy algorithm and the incremental surrogate model with continuously improved accuracy to determine the next sample points at arbitrary locations, and hence, is applicable to the high-dimensional broad flight envelope.

Greedy sampling algorithm aims to seek the maximum of a function, typically the error function between the models of varying fidelity, eliminate this error by including samples at the location corresponding to the maximum values, and hence, improving the accuracy of the model [16]. Therefore, greedy sampling is well-suited for the approximation of high-dimensional models, e.g., using reduced order models, where exhaustive search or manual specification of the search direction can become intractable. For example, Pau-Dubois-Taine and Amsallem [17] presented an adaptive greedy sampling procedure based on a surrogate model of the error indicator to explore the configuration space and identify parameters potentially yielding high error and to update the reduced-order basis for the reduced order model. Bui-Thanh *et al.* [18] developed an adaptive sampling approach for model reduction in high-dimensional parameter input space. They utilized the concept of greedy sampling, and the task of selection of appropriate sample locations is formulated as a sequence of adaptive model-constrained optimization problems in a continuous parameter space without *a priori* specification of a discrete sample set. Hesthaven *et al.* [19] proposed two enhanced greedy sampling algorithms, which are based on assumption of error saturation to reduce workload and the adaptive sparsification of the training set. Both are applicable to high-dimensional modeling problems. It should be noted that while greedy sampling is a heuristic approach, it has been well proven through numerous examples that the underlying mathematical model subject to the high-dimensional inputs can be captured effectively [20, 21].

Surrogate model, also termed metamodel and response surface model, is a numerical method, which mimics the system behavior of physics-based and high-fidelity simulation models through mapping the functional relationship between inputs and outputs in order to achieve significant speedup of computation for a variety of engineering analysis and design optimization [22, 23]. Typical surrogate modeling approaches include multivariate polynomial regression, Kriging, neural network, among others. Falkiewicz *et al.* [24] developed a reduced order aerothermoelastic framework for hypersonic vehicle control simulation, in which the Kriging surrogate model was used to accurately capture the aerodynamic heat flux on a hypersonic control surface within a high-dimensional parameter space, including Mach number, altitude, Angle of Attack, surface deformation and surface temperature. In a similar MOR development for hypersonic aerothermoelasticity, Crowell and McNamara [25] employed the second-order polynomial to approximate the surface temperature of a wing structure in two directions and the Kriging and/or proper orthogonal decomposition methods for modeling the surface heat fluxes. The author's group [26] developed a holistic framework that integrates aerodynamics, structure dynamics, and active structural control for closed-up ASE analysis, in which surrogate modeling methods including autoregressive moving average and artificial neural network were employed for generalized aerodynamic force prediction under various structure deformation and flight conditions. The aeroelasticity analysis of the high-Reynolds-number aerostructural dynamics configuration was also performed with a nonlinear surrogate model approach RBF-NN [27].

In contrast to previous efforts, the foremost goal of the present effort is to combine the greedy sampling algorithm and the incremental surrogate modeling for selecting the most critical grid points and tailoring the database of the LPV ASE model for flexible aircraft. There are several points of note for the current development: (1) although our method follows the greedy sampling algorithm to sequentially add grid points into the sampled pool, the aim is to remove the redundant LTI models—as many as possible—without compromising modeling

accuracy; (2) The original database constructed by the high-fidelity physics-based modeling with dense grid distribution already exists, which is naturally amenable to greedy sampling that selects from discrete sample values. On the contrary, the model-constrained optimization to determine samples in the continuous domain [18] is less relevant in our case; (3) while Kriging will be used as the incremental surrogate model in this paper, the methodology and framework formulated in this paper is generally applicable to any surrogate model, such as polynomial regression and neural network; and (4) other adaptive sampling techniques, such as those based on the estimate of prediction error, and weighted measures of exploration and exploitation can also be utilized in lieu of the greedy sampling algorithm [22, 28]. However, since the original database and the initial sampled pool are already available for selection, greedy sampling is especially appropriate for database tailoring.

## II. Linear Parameter-Varying Aeroservoelastic Models of Aircraft

Linear parameter-varying (LPV) models are state-space models whose state-space descriptions are functions of time-varying parameters, i.e.,

$$\begin{bmatrix} \dot{x} \\ y \end{bmatrix} = \begin{bmatrix} A(\rho) & B(\rho) \\ C(\rho) & D(\rho) \end{bmatrix} \begin{bmatrix} x(t) \\ u(t) \end{bmatrix} \quad (1)$$

where  $A(\rho)$  is the state matrix,  $B(\rho)$  is the input matrix,  $C(\rho)$  is the output state matrix,  $D(\rho)$  is the input transition matrix,  $\rho \in \mathcal{R}^{np}$  is a vector of measurable parameters (such as altitude, Mach, fuel weight, etc.),  $u$  and  $y$  are, respectively the vector of the control input and measurement. In this paper, the representation of the parameter dependence in LPV models is fulfilled through linearization on a gridded domain to agree with the ASE models in the original database. The gridded domain LPV is illustrated in Figure 1, in which the nonlinear dynamics in the ASE problem of the aircraft is treated as its linearization around various flight operating points (also termed grid points or parameter locations hereafter). A set of Linear Time Invariant (LTI) state space models are first constructed at stencil grid points with a high resolution with a multi-dimensional parameter space, leading to a model database with dense distribution of grid points.

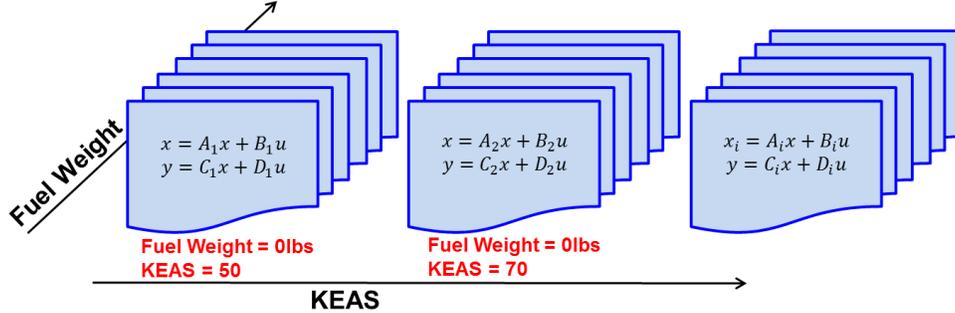


Figure 1. Linear parameter varying (LPV) formulation of the aeroservoelastic (ASE) models of aircraft

## III. Greedy Sampling and Surrogate Modeling-based Database Tailoring

Figure 2 illustrates our greedy sampling and surrogate modeling-based methodology for tailoring the database of LPV ASE models. As discussed above, the models in the original database are constructed at the stencil grid points in the multidimensional flight parameter space. The process starts with generating a few initial samples  $s_1$  given a specified sample number, and an empty bin  $\mathbf{S}$  of sampled/selected grid points and a full bin  $\mathbf{U}$  of unsampled grid points (that initially contains all the grid points). The initial sample  $s_1$  is then added into the sampled bin  $\mathbf{S}$ , and meanwhile removed out of the unsampled bin  $\mathbf{U}$ . The entries in  $\tilde{A}$ ,  $\tilde{B}$ ,  $\tilde{C}$ , and  $\tilde{D}$  matrices of sampled points in  $\mathbf{S}$  are then used to construct a data-driven surrogate model  $\hat{G}(s_j) = (\hat{A}, \hat{B}, \hat{C}, \hat{D})$  as a function  $s_j$ , where  $s_j \in \mathbf{U}$  is arbitrary grid point in the unsampled bin  $\mathbf{U}$ . Note that unless annotated, the tilde symbol  $\tilde{\cdot}$  denotes the quantities at the unsampled grid points, i.e., set  $\mathbf{U}$ ; while the hat symbol  $\hat{\cdot}$  denotes the quantities at the sampled grid points, i.e., set  $\mathbf{S}$ .

In the present effort, the Kriging surrogate model is used (see Section A). The surrogate model interpolation is then undertaken to predict the state space model  $\hat{G}(s_j)$  and its system matrix entries at unsampled grid points  $s_j \in \mathbf{U}$ . The quantitative error  $e$  between the original model  $G$  and the interpolated model  $\hat{G}(s_j)$  will be computed and compared at all unsampled grid points in  $\mathbf{U}$ . If the maximum of  $e$  is below a specified tolerance  $\varepsilon$ , the program will terminate yielding a sparse model database that satisfies the accuracy requirement. Otherwise the grid point  $s_k \in \mathbf{U}$

that carries the worst error relative to the original models will be selected as a new sample to be added in the sampled bin  $S$  and removed out of  $U$ . With enriched samples in  $S$ , the surrogate model will be incrementally updated to further resolve complex response surface topology in in the system. The process will be repeated till the accuracy requirement is met. Accordingly, two components are developed and integrated in our framework to fulfill the process, namely, surrogate modeling and greedy sampling, and are presented in the following sections:

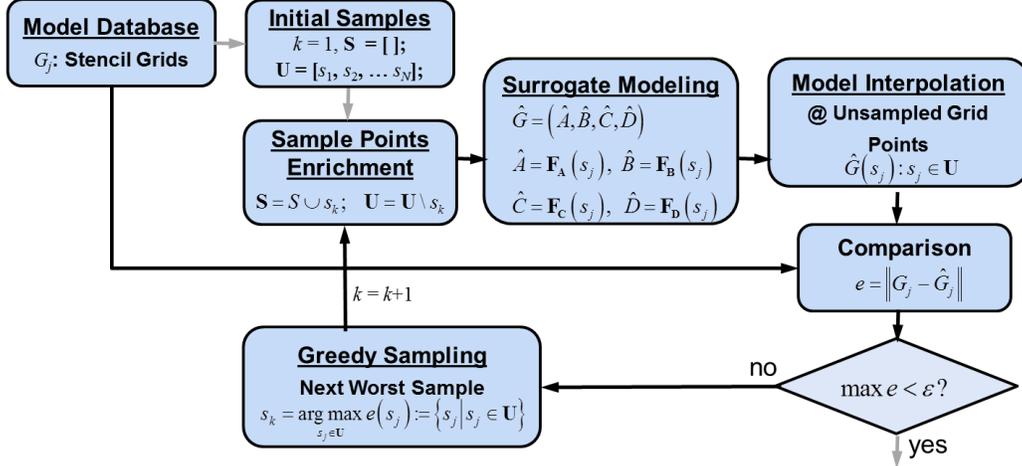


Figure 2. Organization of greedy sampling and surrogate modeling-based tailoring of model database

### A Surrogate Modeling

There are a variety of methods to construct data-driven surrogate models, including Kriging, polynomial regression, neural network, radial basis function, among many others. Kriging is one of the interpolation approaches, in which the values to be interpolated are dictated by a Gaussian process and prior covariance. In contrast to others, Kriging has proven well-suited for approximation of nonlinear functions without *a priori* assumptions on the form of the function [29]. Recent studies also confirm that Kriging yields relatively accurate results when the training data for model construction is sparse [30]. Therefore in this paper the Kriging method is selected to build the surrogate models. However it should be pointed out that work flow outlined above is generally extendable to any surrogate modeling method.

In Kriging, the surrogate model  $F_l$  that establish the mapping relationship between  $s_j$  and the entries in system matrices  $\{\hat{A}, \hat{B}, \hat{C}, \hat{D}\}$  is given by

$$F_l(s_j) = \bar{f}_l(s_j) + \mathbf{f}_l(s_j) \quad (2)$$

where  $F_l(s_j)$  is the Kriging model to predict the elements in system matrices at a flight condition  $s_j \in U$  that has not been sampled, and subscript  $l = \{A, B, C, D\}$  to denote the input, state, output and direct transition matrix.  $\bar{f}_l(s_j)$  is an assumed global trend function (usually a low-order polynomial) of  $F_l(s_j)$ , and  $\mathbf{f}_l(s_j)$  takes into account the local deviations and is assumed to be a random function following a stochastic process.  $\mathbf{f}_l(s_j)$  enables the Kriging interpolator to effectively capture the nonlinear behavior regardless of the exact function's form. The covariance matrix used to characterize the correlation between two samples points is given by

$$\text{Cov}[\mathbf{f}_l(s_k), \mathbf{f}_l(s_m)] = \sigma^2 \mathbf{R}_{krig}(s_k, s_m) \quad (3)$$

where  $\sigma^2$  is the process variance above, and  $\mathbf{R}_{krig}$  is the matrix to correlate two sample points  $s_k$  and  $s_m$  and can be calculated using a user defined spatial correlation function (SCF). Herein, the Gaussian correlation function is used

$$(\mathbf{R}_{krig})_{p,m} = R_{krig}(\tilde{s}_p, \tilde{s}_m) = \prod_{n=1}^{N_d} \mathfrak{R}(\theta, d_n), \quad \mathfrak{R}(\theta, d_n) = \exp(-\theta_n d_n^2) \quad \text{and} \quad d_n = \tilde{s}_{n,p} - \tilde{s}_{n,m} \quad (4)$$

where  $\theta_k$  are the correlation parameters that need to be determined by fitting the data at sample points based on the maximum likelihood estimation. Again symbol  $\sim$  denotes quantities of the already selected grid points in  $S$ ; and  $N_d$  is the dimension of the flight parameter space, and herein,  $N_d = 2$  for the air speed and fuel weight.

Once Eqs. (3) and (4) are available, the system matrices of  $\hat{G}$  at unsampled grid points can be computed using Eq. (2). Take the state matrix  $\hat{A}$  as an example, it is given by [29, 31]

$$\hat{A} = \mathbf{F}_A(s) = \mathbf{g}^T \mathbf{w} + \mathbf{r}_{krq}^T(s) (\mathbf{R}_{krq})^{-1} (\tilde{A} - \tilde{\mathbf{H}} \mathbf{w}) \quad \text{and} \quad \mathbf{w} = \left( \tilde{\mathbf{H}}^T (\mathbf{R}_{krq})^{-1} \tilde{\mathbf{H}} \right)^{-1} \tilde{\mathbf{H}}^T (\mathbf{R}_{krq})^{-1} \tilde{A} \quad (5)$$

where  $\mathbf{g}^T$  consists of the basis functions/regressors of the low-order polynomial for the global trend function; and  $\mathbf{w}$  denotes the regression coefficient and is evaluated using the selected grid points.  $\tilde{\mathbf{H}}$  is the design matrix for regression constructed using the selected grid points in  $\mathbf{S}$  and  $k$  is the number of the selected grid points in  $\mathbf{S}$  in the current iteration.  $\mathbf{r}_{krq}(s)$  in Eq. (5) is given by

$$\mathbf{r}_{krq}(s_j) = \left[ R_{krq}(s_j, \tilde{s}_1) \quad R_{krq}(s_j, \tilde{s}_2) \quad R_{krq}(s_j, \tilde{s}_k) \dots R_{krq}(s_j, \tilde{s}_{N_s}) \right]^T \quad (6)$$

where  $s_j \in \mathbf{U}$  and  $\tilde{s}_m \in \mathbf{S}$  are, respectively, the grid points in the unsampled and sampled bins.

In this study, the Design and Analysis of Computer Experiments (DACE) [31] toolbox in MATLAB that implements the universal Kriging is used. In the DACE toolbox the trend function is a low-order polynomial regression function that can be either constant, linear, or quadratic, and the regression coefficients are obtained by the generalized least-squares method.

## B Greedy Sampling

The underlying principle of the greedy sampling is to sequentially add in the sampled bin a new sample point at the location in the parameter space where the error  $e$  between the interpolated surrogate model  $\hat{G}$  and the physics-based benchmark model  $G$  is maximum [19], which is given by

$$s_k = \arg \max_{s_j \in \mathbf{U}} e(s_j) := \{s_j \mid s_j \in \mathbf{U}\} \quad (7)$$

Note that hereafter the models on grid points constructed using the physics-based high-fidelity simulation tool, i.e.,  $G$  in the original database and the surrogate models built with  $G$  on all the grid points are termed ‘‘benchmark model’’ and ‘‘benchmark surrogate model’’, respectively. The fidelity of the surrogate models can be increasingly improved as a result of continuous sample point enrichment and refinement of the response surface topology. This step is repeated until the updated surrogate model yields excellent agreement with the benchmark models at the unsampled points  $\mathbf{U}$ . Despite a heuristic approach, it has been well established in the literature that the underlying mathematical model and observable outputs can be captured very well [16, 18, 20].

In contrast to the conventional greedy sampling that requires *a priori* trial sample set typically selected randomly, our initial sample set  $\mathbf{U}$  already exists and contains densely placed stencil grid points in the flight parameters of the original database. Moreover our error  $e$  indicates the mismatch in the frequency response along all input-output channels, and is mathematically given by

$$e = \left| \frac{G - \hat{G}}{G} \right| \quad (8)$$

That is, the discrepancy between the benchmark and the interpolated model is numerically computed and its absolute value is accumulated at pre-determined frequency values in the targeted range along all input-output channels. We have also tested other error indicators, such as  $H_\infty$  norm [32], and it was found that Eq. (8) allows best representation of the difference and consistent results.

## IV. Results and Discussion

The LTI state-space models of the aircraft airframe were provided by NASA/AFRC. They were developed using the generalized mass, stiffness, and aerodynamic matrices obtained by MSC/Nastran [33] and ZAERO [34]. There are 10 control surfaces on the vehicle, five on each wing; and 2 throttle controls for engine dynamics. The five actuators on the left wing are labeled as BFL, WF1L, WF2L, WF3L, and WF4L starting from the inner body to the outer wing tip. Likewise, the actuators on the right wing are labeled as BFR, WF1R, WF2R, WF3R, and WF4R based on the same convention above. The rigid-body state sensors (IMU-MIDG) are located around the center of the vehicle, while the six accelerometer locations are, respectively, placed at the front of the vehicle (ASENSR100), at the rear (ASENSR1000), at the leading and trailing edge of the left wing (ASENSR400 and ASENSR600), and of the right wing (ASENSR1100 and ASENSR1300).

A set of 442 models were generated at  $M = 0.16$  on grid points of a 2D parameter space across the flight envelope. The two parameters are KEAS (‘‘knots equivalent air speed’’) and fuel weight, which, respectively, range from 50 KEAS to 175 KEAS in 5 KEAS increments and from 0 lb to 80 lb in 5 lb increments (corresponding to the 442 models). The model at each grid point has 247 states, including 56 states corresponding to the 2<sup>nd</sup>-order sensors

(28 sensors in total), 12 rigid body states, 25 elastic structural modes and 25 derivatives (modal velocity), 93 aerodynamic lag states, and 36 states for the third order actuators (10 control surfaces and 2 engine throttles). According to the V-g and V-f plots of the baseline model [35], the normalized flutter frequencies for SBFF (symmetric body freedom flutter), SWBTF (symmetric wing bending torsion flutter), and AWBTF (anti-symmetric wing bending torsion flutter) modes are, respectively, at 1, 3.68, and 3.912 (all the flutter frequencies are normalized by the one for SBFF due to ITAR requirement). The target normalized frequency range  $\omega$  is determined to be  $0.01 < \omega < 5.37$  to ensure full coverage of the instability of interest and system response. The detailed model information and the physical meaning of the states is given in [9].

The methodology of database tailoring described in Section III was then applied to the ASE model of the aircraft. Specifically, the process starts with 10 initial grid points selected by a variant of the Latin Hypercube Sampling (LHS), that is, the continuous parameters selected by LHS design are rounded to the closest discrete grid point in the flight envelop. The entries in model matrices  $\hat{G} = \{ \hat{A}, \hat{B}, \hat{C}, \hat{D} \}$  at the selected grid points are then used to construct the surrogate model. The predicted model matrix entries by use of the surrogate model are then compared against the benchmark model at those unsampled/unselected grid points, among which the one yielding the largest relative error  $e$  is then picked as the next sample to enrich the sample pool. The relative error  $e$  is numerically calculated using Eq. (8) at 500 predetermined log-spaced frequency values in targeted range among all input-output channels. Because of the paper limit, only 2 inputs and 2 outputs out of the 12 inputs and 13 outputs are selected for visualization purpose. Their names and corresponding sensor and actuator names are listed in Table 1.

**Table 1: Input and output channel names and their corresponding actuator and sensor names for the model**

Input No.	Input Name	Actuator Name	Output No.	Output Name	Sensor Name
1	BFL C FCS DEG	BFL	1	Q MIDG DPS	$Q$
2	WF4L C FCD DEG	WF4L	2	LOFZ VMC 200 G	ASESNSR 400

### A Prediction with Increasing Sampling Points

We first evaluate the effects of increasing sampled/selected grid points on the surrogate model and interpolation accuracy. In this study, the surrogate model with a trend function of a 2<sup>nd</sup>-order polynomial, i.e.,  $\mathbf{g}^T \mathbf{w}$  term in Eq. (5). Figure 3a-c illustrates the distribution of the first 50, 100, and 150 grid points selected by the greedy sampling algorithm. Figure 3a shows the distribution of the first 50 grid points that are more important in terms of capturing the dependence of the system behavior on flight parameters. Although sparse, they are mostly located at the top and the left boundaries of the flight envelop, corresponding to the large fuel weights and low air speed. Figure 3b illustrates the distribution of the 100 selected/sampled grid points contributing most to the error reduction. Again selected grid points are mostly placed at the region of low airspeed and large fuel weights, where the flight dynamics of aircraft is more susceptible to the variation of the flight parameters. Figure 3c illustrates the location of the first 150 selected/sampled grid points, and it is also found that the terminal flight conditions, i.e., the four boundaries of the flight envelop, are more densely sampled because the sample points are only available from one side in contrast to the interior region. Within the interior region, in general the distribution of the selected grid points is sparser for the low fuel weight and high air speed, i.e., the bottom right region.

Figure 3d illustrates the decay of the worst relative error  $e$  as grid points are added into the sampled pool for surrogate modeling. It is clear that the worst relative error  $e$  decreases dramatically and almost exponentially, and falls below 1% even with 100 grid points selected. It is also interesting to note that addition of new sample points may alter the response surface among various input-output channels predicted by the surrogate model and their weights contributing to the overall errors, resulting in several jumps in the error. The minor oscillation of the relative error is in particular obvious when  $e$  is already very low (e.g.,  $< 1\%$ ) and a large number of grid points are selected ( $> 150$  points) because a slight change in predication essentially can be amplified, which is in distinct contrast to beginning of the process when only 50 sampled points are taken. Nonetheless a general trend of continuous error reduction is explicit in Figure 3d.

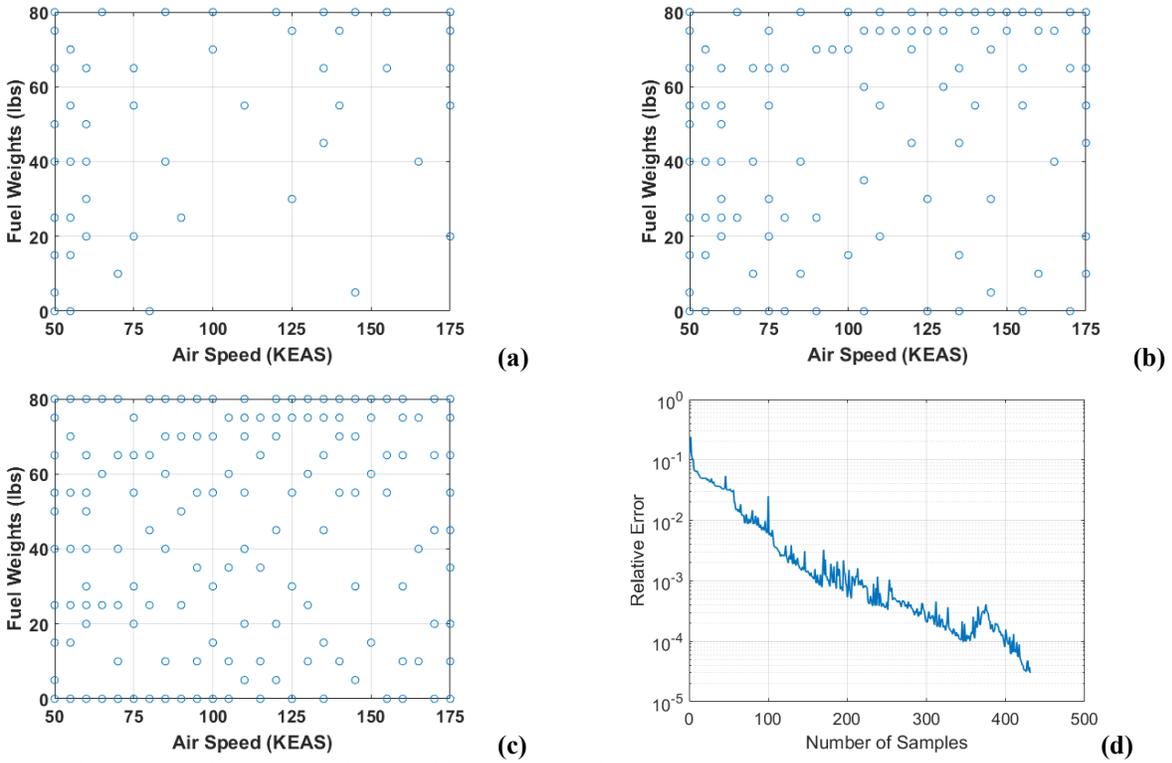


Figure 3. Distribution of (a) 50, (b) 100, and (c) 150 selected samples. (d) Relative error  $e$  vs. number of samples.

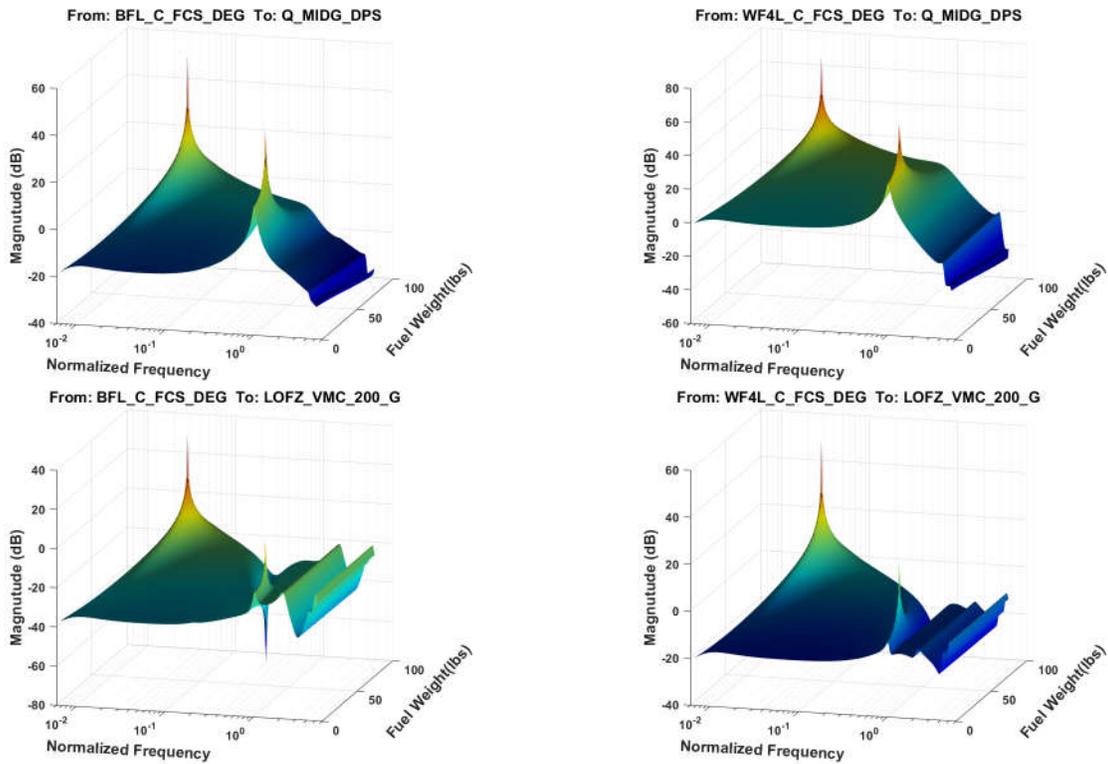


Figure 4. Bode diagram as a function of fuel weight at a fixed speed of 112 KEAS predicted by the benchmark surrogate model constructed using all grid points.

In the following, the frequency responses and model interpolation at non-grid points obtained from three surrogate models are quantitatively compared. In the first surrogate model, i.e., the benchmark surrogate model, the system matrix entries at all the grid points (442 models on the stencil grid points) are used, and its interpolation results are used as the benchmark. In the 2<sup>nd</sup> and the 3<sup>rd</sup> surrogate model, only the first 50 and 150 grid points are selected and used. Figure 4 illustrates the magnitude of Bode diagram from the two inputs (BFL and W4FL) to the two outputs ( $q$  and ASESNSR 400) listed in Table 1. It is predicted by the benchmark model as a function of the fuel weight (lbs) while the airspeed is kept constant at 112 KEAS. The interpolation is carried out with a resolution of 1 lbs (including non-grid points) for enhanced visualization. Figure 5 illustrates similar results obtained through the 2<sup>nd</sup> surrogate model that only uses the first 50 selected grid points for model construction. Although the responses in both figures exhibit similar profiles, the difference in quantitative values as manifested by the color intensity between them is evident. Significant mismatch is observed at the regime of low frequency and low fuel weight. The interpolation accuracy is markedly improved by incorporating benchmark models  $G$  at more grid points. Figure 6 illustrates the frequency responses predicted by the surrogate model constructed using the first 150 grid points, which is almost indistinguishable from the benchmark surrogate model case (using all grid points) in Figure 4.

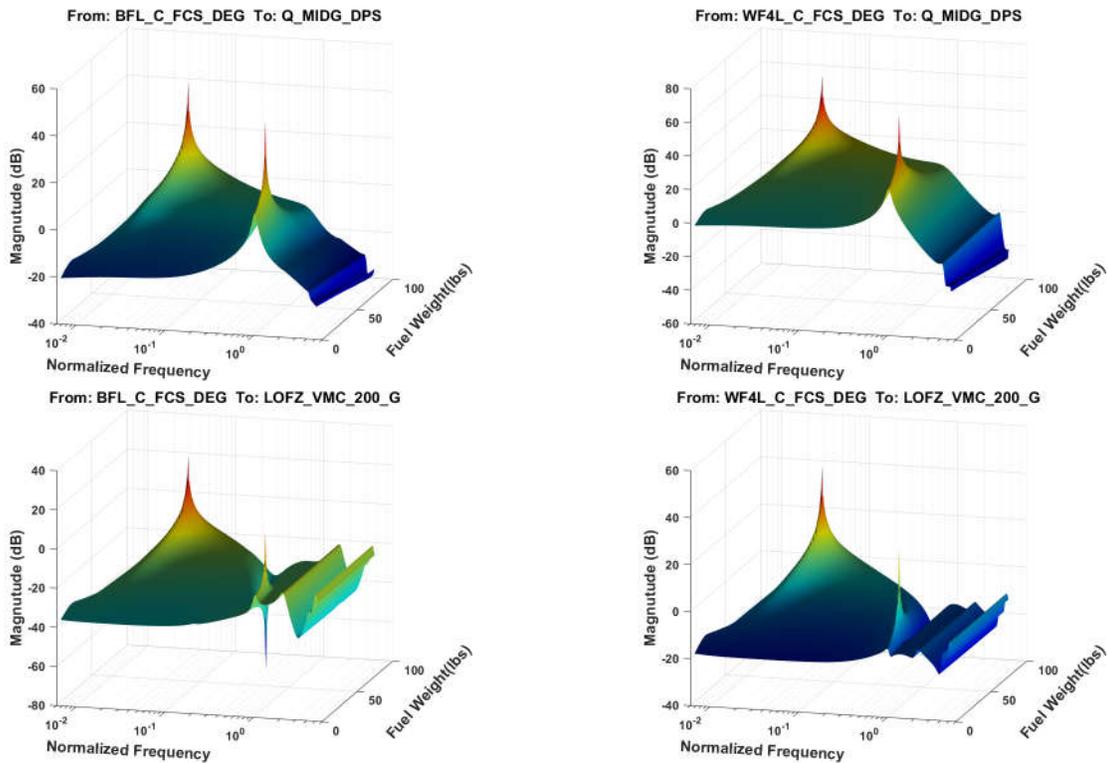


Figure 5. Bode diagram as a function of fuel weight at a fixed speed of 112 KEAS predicted by the surrogate model constructed using the first 50 selected grid points.

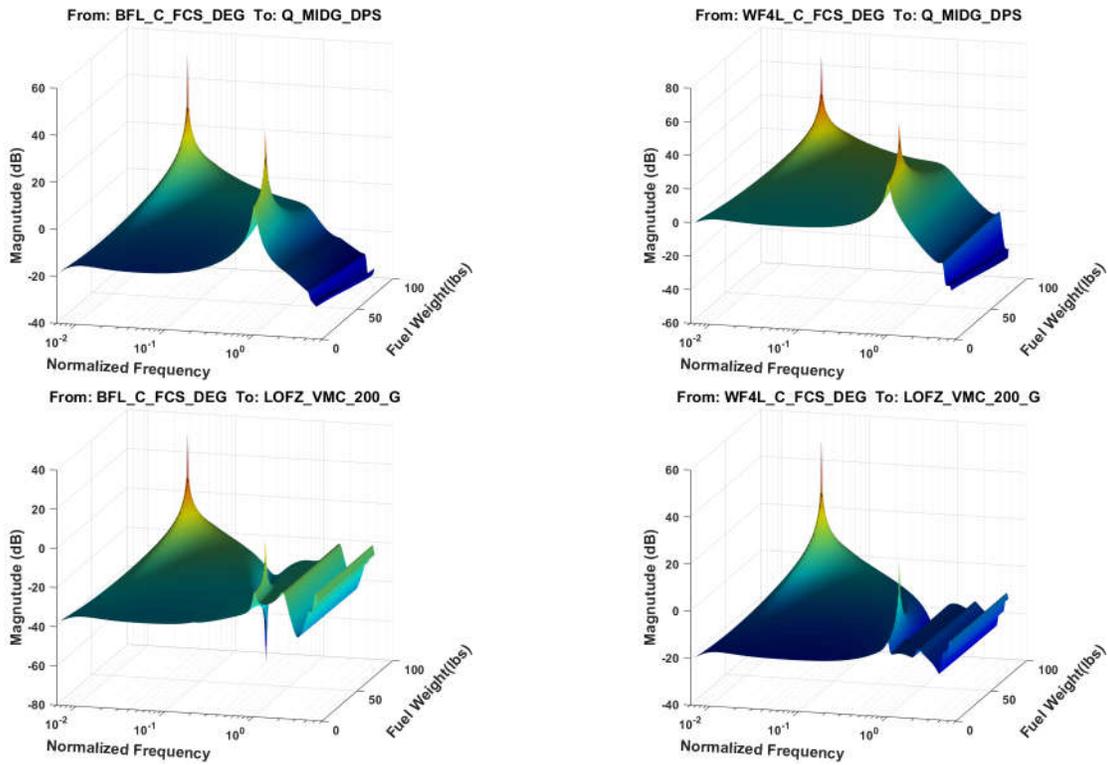


Figure 6. Bode diagram as a function of fuel weight at a fixed speed of 100 KEAS predicted by the surrogate model using the first 150 selected grid points.

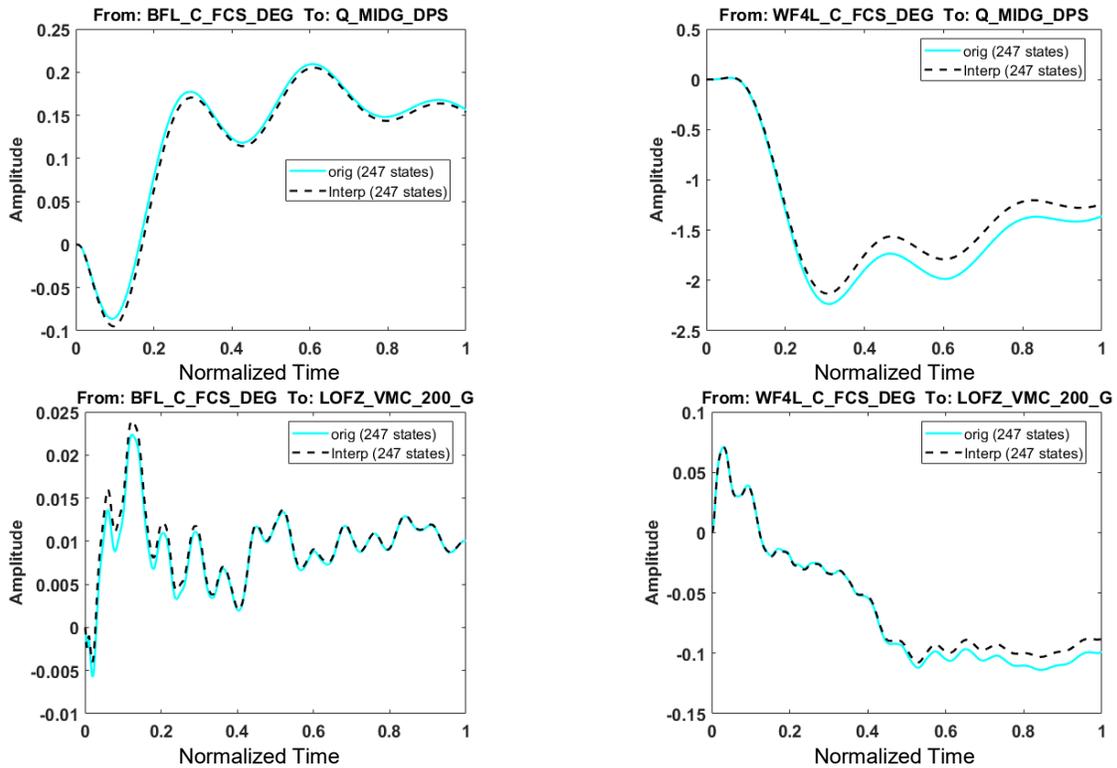
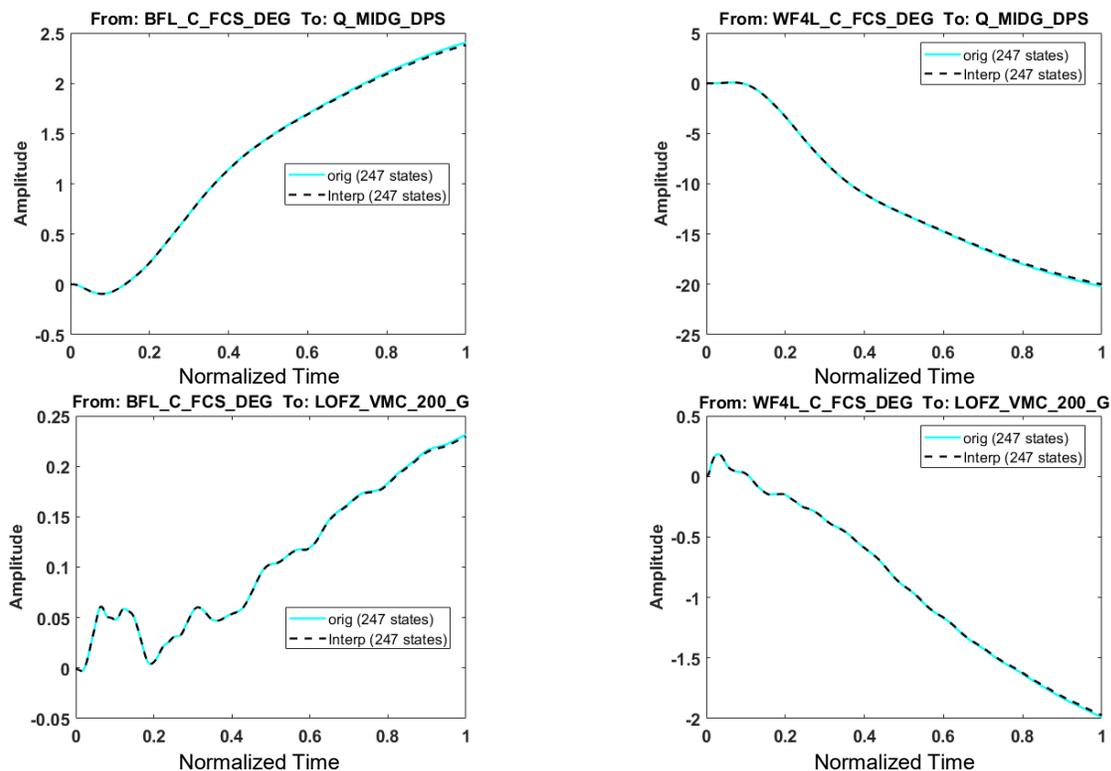


Figure 7. Step response plot of benchmark surrogate model (constructed using all grid points) vs. the surrogate model using the first 50 selected grid points



**Figure 8. Step response plot of benchmark surrogate model (constructed using all grid points) vs. the surrogate model using the first 150 selected grid points**

The observations in the frequency domain are also reflected in the time-domain as shown in Figure 7 and Figure 8. Figure 7 shows the comparison in the step response in the normalized time domain among the input-output channels predicted by two surrogate models constructed, respectively, using the first 50 grid points (labeled as “interp”) and using all the grid points (labeled as “original”, i.e., benchmark surrogate model in Figure 4). The comparison is performed at the next worst-error grid point, i.e., the 51th grid point in Figure 3d. Thus, the “interp” curve in Figure 7 is predicted through Kriging interpolation of the surrogate model using 50 grid points. Both models match very well at the small-time scale and then deviate from each other at the large-scale, which is consistent with Figure 5 where the surrogate model disagrees distinctly with the benchmark model at the low frequency regime. Figure 8 illustrates the results for the surrogate model constructed using 150 grid points, which shows excellent accord with the benchmark surrogate model without any noticeable discrepancy. Note that in this case the comparison is made at the 151th grid point, i.e., the one of the worst error for the former.

A striking finding is that eventually LPV ASE models at only 150 grid points are adequate for surrogate modeling and interpolation to construct this aircraft model at arbitrary locations within the flight envelope. Given the original database consisting of 442 models, it amounts to almost 66% decrease in the model number, and potential for model order reduction, and control synthesis and implementation. This verified that despite a heuristic approach, the greedy algorithm can depict very well the underlying mathematical model and observable outputs.

## B Effect of Trend Function and Initial Sampling

We also carried out a study to investigate the effect of trend function and initial sampling on the ASE model database tailoring. Specifically, only the linear terms (i.e., the first-order polynomial) of the flight parameters are used in the trend function, and a different set of 10 samples generated by the LHS variant initializes the process.

Figure 9a-c depicts the distribution of the first 50, 100, and 150 grid points selected by the greedy sampling algorithm using the new linear trend function and initial sampling. It is virtually the same as those in Figure 3 although the exact locations of the selected points are different. The first 50 points mostly stay around the vicinity of the top and the left boundaries of the flight envelop. When 150 grid points are selected for surrogate modeling, a lot of them are located at the four boundaries and the top left corner of the interior region as shown in Figure 9c, which again agrees with Figure 3c. These consistent observations confirm that the non-uniform distribution of the grid

points is attributed to the flight dynamics of the aircraft model, i.e., the dependence of the ASE response on the flight parameters, rather than the random initial sampling or use of different trend functions. Figure 9d illustrates the worst relative error  $e$  vs. the number of grid points added into the sampled pool for surrogate modeling. It is also within our expectation that the relative error decays almost exponentially and at the same rate as that in Figure 3d.

Similarly the two surrogate models built with all grid points and only 50 grid points are quantitatively compared in terms of the frequency responses and model interpolation at non-grid points. The results of the model constructed using 150 selected grid points are not shown herein as they are almost exactly the same as the benchmark surrogate model. Figure 10 illustrates the magnitude of Bode diagram from the two inputs (BFL and W4FL) to the two outputs ( $q$  and ASESNSR 400) predicted by the benchmark surrogate model. Different from Figure 4, in this study the air speed is varied from 50 to 175 KEAS while the fuel weight is kept constant at 40 lbs. The interpolation is carried out with a resolution of 2 KEAS that also include non-grid points. Figure 11 exhibits the frequency response predicted by the surrogate model that is constructed using only the first 50 selected grid points. Although its overall contour resembles that in Figure 10, notable difference in quantitative values, peak magnitudes, and local variations is clearly observe, in particular, at the regime of low frequency and low air speed. For example, the wrinkle behavior between 50 and 100 KEAS in the graphs on the left of Figure 10 are not captured at all. Adding more grid points, e.g., 150 grids points dramatically improve the interpolation accuracy and capability to fully resolve the response surface topology (data is not shown because it is almost identical to Figure 10).

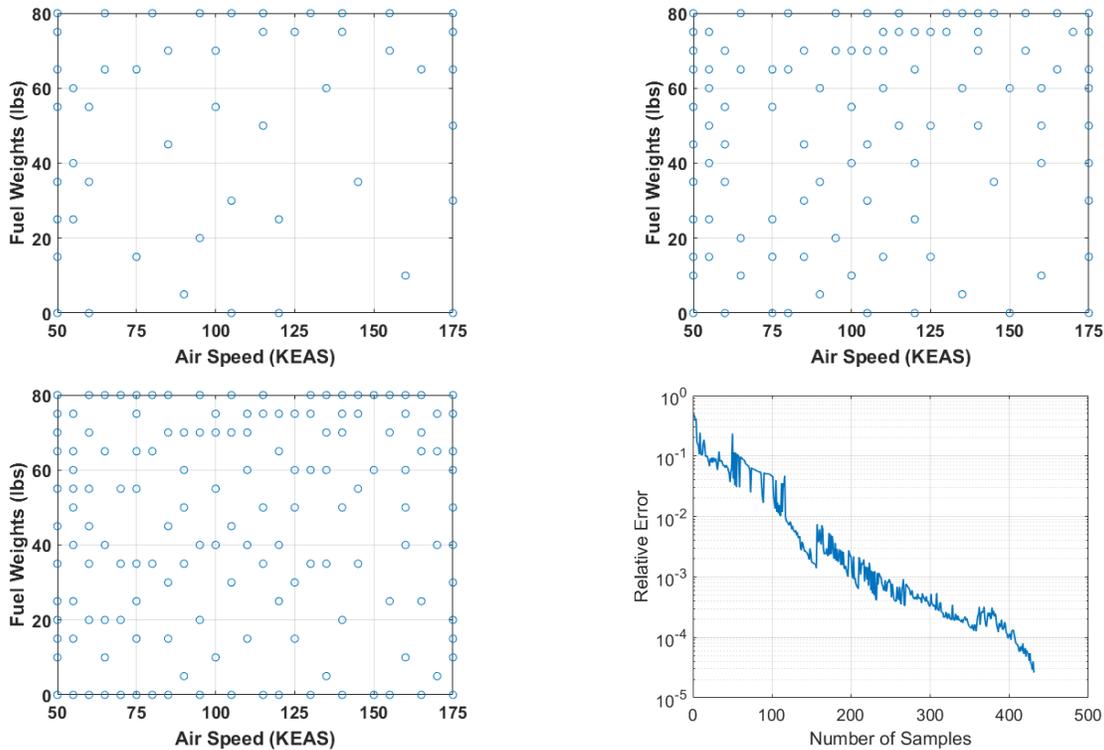


Figure 9. Distribution of (a) 50, (b) 100, and (c) 150 selected samples. (d) Relative error  $e$  vs. number of samples.

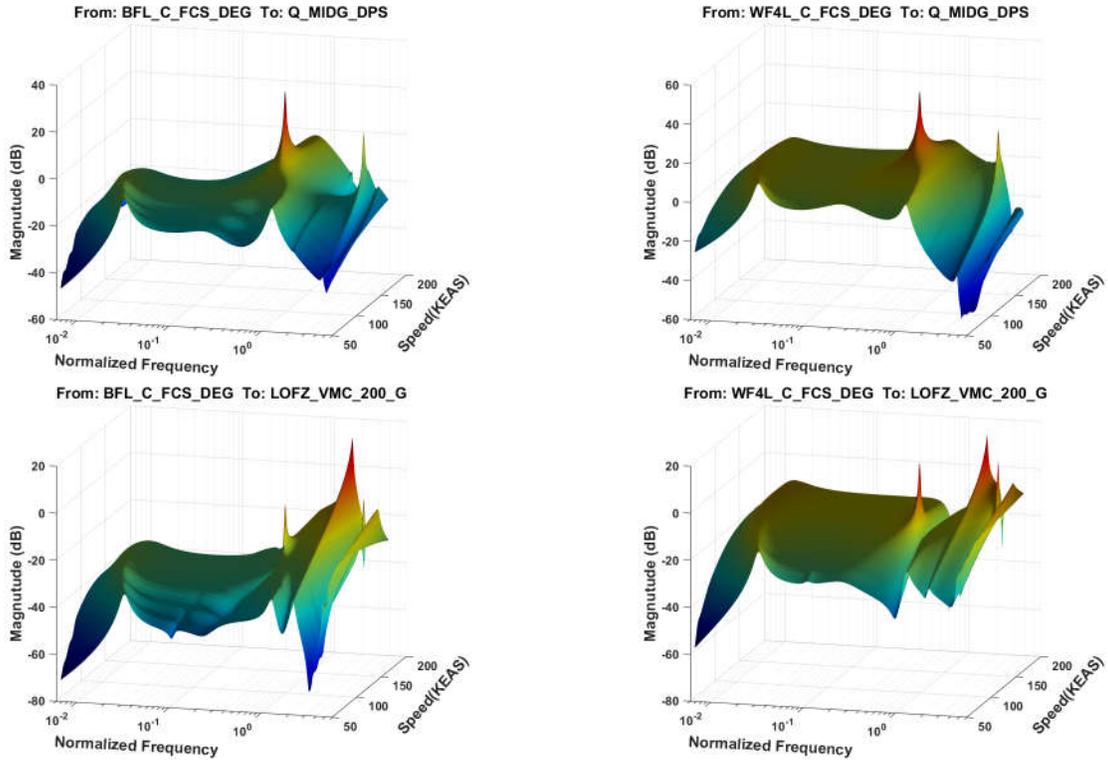


Figure 10. Bode diagram as a function of air speed at a fixed fuel weight of 40 lb predicted by the benchmark surrogate model using all grid points.

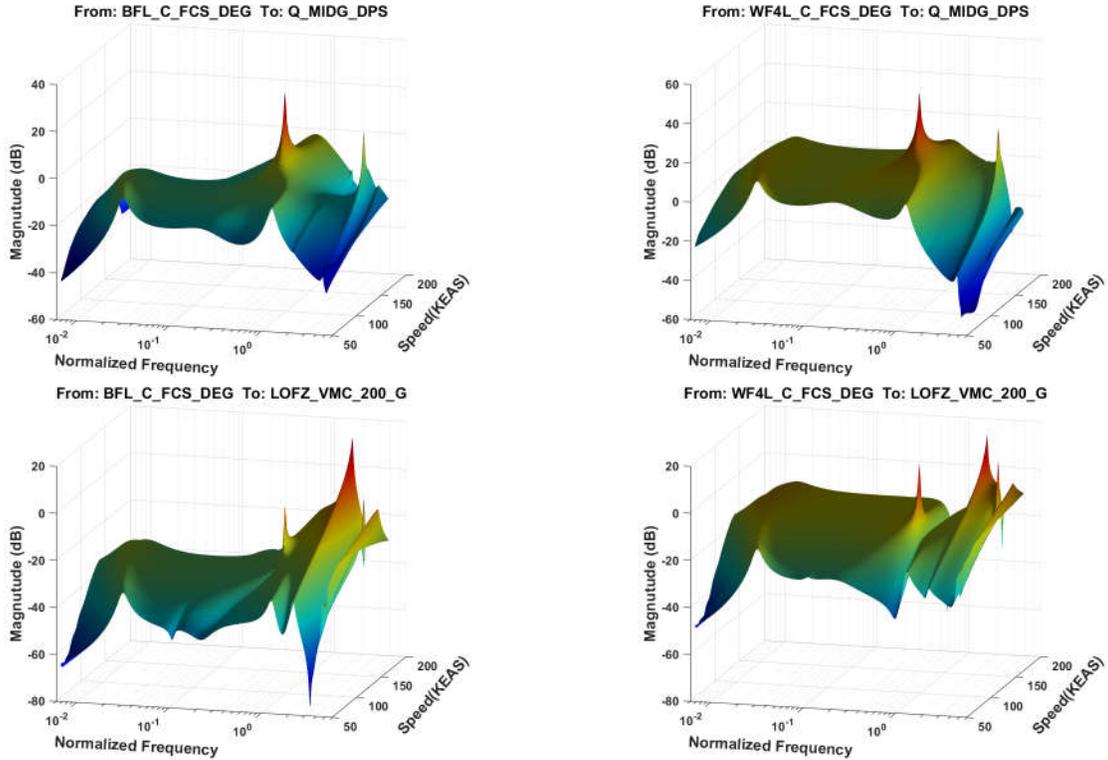
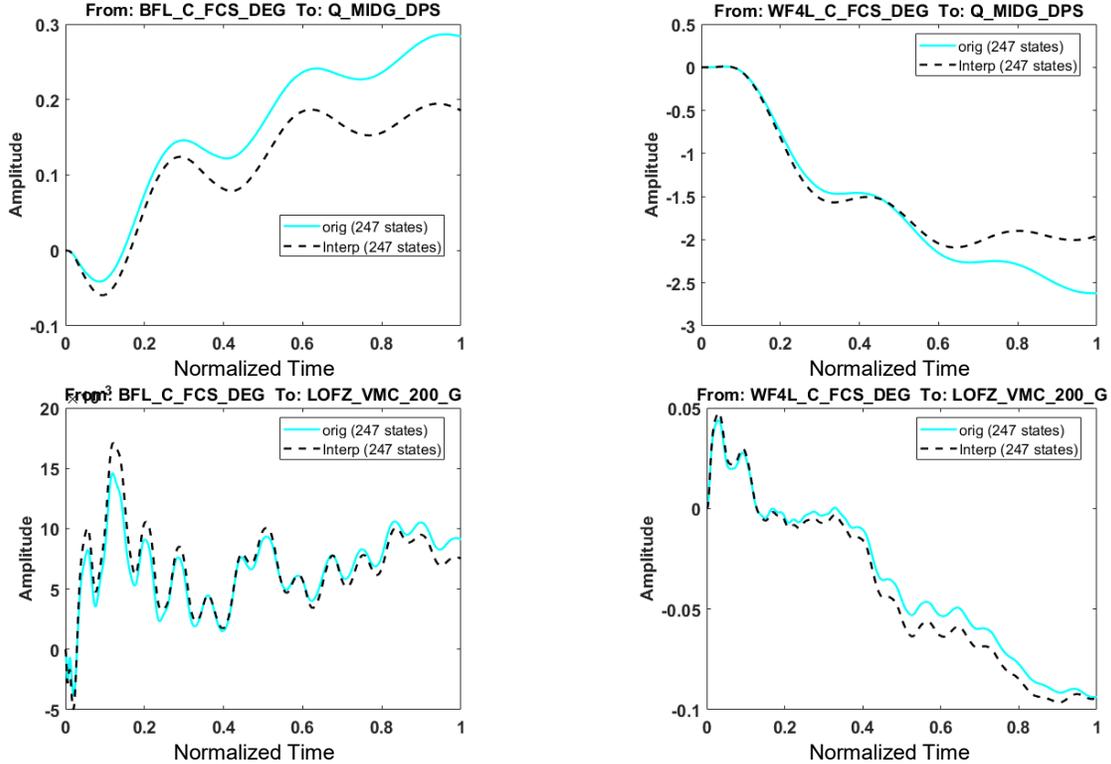


Figure 11. Bode diagram as a function of air speed at a fixed fuel weight of 40 lb predicted by the surrogate model constructed using the first 50 selected grid points.



**Figure 12. Step response plot of benchmark model (constructed using all grid points) vs. the surrogate model using the first 50 selected grid points**

The discrepancy in the frequency domain also translates to the time domain as shown in Figure 12, which shows the comparison in the step response (in the normalized time domain) among the input-output channels predicted by two surrogate models constructed above. Again both are compared at the next worst-error grid point, i.e., the 51th grid point in Figure 9d, and the labels of “interp” and “orig” curves are annotated above. Both models show good agreement at the beginning and then depart from each other at the large time scale. Similarly incorporation of additional grid points could drastically enhance the prediction of the time-domain response (data not shown).

## V. Conclusion

An automated data analysis and modeling framework has been presented for tailoring LPV ASE systems for flexible aircrafts. The process includes two key steps, i.e., incremental surrogate model construction and sample selection. In the former, the Kriging-based surrogate model was constructed using ASE models at a fraction of selected grid points within the parameter space of the original model database, and then utilized to predict the models at arbitrary locations. In the latter, the error of surrogate ASE model prediction relative to the benchmark surrogate model in the frequency response among all input-output channels is then calculated at unsampled grid points, and the greedy sampling algorithm is proposed to select the one with the worst relative error. The process is repeated to keep updating the surrogate model till a pre-determined tolerance or iteration budget is met. This method is essentially to perform an adaptive space-filling sampling on an existing densely-gridded model database, leading to compact database structure and salient applicability to high-dimensional flight parameter space.

The database tailoring methodology was verified using the NASA’s ASE model database of a flexible aircraft, which included 442 state-space models (247 states each) built on a stencil grid in 2D parameter space of air speed and fuel weight. Our investigations show that the number of the models could be reduced down to 150 by the proposed approach, equivalent to 67% reduction. Even so the ASE models obtained through Kriging interpolation at the unsampled grid points match the benchmark model in the original database constructed directly from the physics-based numerical software with the worst relative error far below 1%. The frequency-domain response of the interpolated ASE model shows smooth variation with the flight parameters, and excellent agreement between the interpolated model and the benchmark model is observed. It is also found that the selected grid points exhibit non-uniform distribution in the flight envelop, implying the distinctly different dynamic behavior and its dependence on

flight parameters. Despite a heuristic approach, the greedy algorithm in general can quickly identify the key grid points and reduce the errors for surrogate model construction, and the underlying response surface of the system matrices in the original model database can be resolved very well.

In addition the case study to interrogate the effect of the trend function in the Kriging model and initial sampling is also undertaken. For our ASE model, both of them have minor influence on the sample distribution, e.g., sparse distribution at the high air speed and low fuel weight, and the accuracy of interpolated model is almost the same.

The future work will focus on integrating the database tailoring approach with the ASE model construction process, i.e., generating a proper number of the model at the first place, model order reduction framework [8, 9], and control synthesis to mitigate the modeling and design workload and develop more robust flight control. The use of the  $\sigma$ -shifted  $H_2$  norm [14] to speed up the tailoring process will also be investigated.

### Acknowledgments

This research is sponsored by NASA Armstrong Flight Research Center under contract NX15CD03C.

### References

1. Sheta, E.F., et al., *Computational and experimental investigation of limit cycle oscillations of nonlinear aeroelastic systems*. Journal of Aircraft, 2002. **39**(1): p. 133-141.
2. Silva, W., et al. *Development of Aeroservoelastic Analytical Models and Gust Load Alleviation Control Laws of a SensorCraft Wind- Tunnel Model Using Measured Data*. in *47th AIAA/ASME/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. 2006.
3. Hjartarson, A., P.J. Seiler, and G.J. Balas. *LPV Aeroservoelastic Control using the LPVTools Toolbox*. in *AIAA Atmospheric Flight Mechanics (AFM) Conference*. 2013.
4. Moreno, C.P., P.J. Seiler, and G.J. Balas, *Model Reduction for Aeroservoelastic Systems*. Journal of Aircraft, 2014. **51**(1): p. 280-290.
5. Poussot-Vassal, C. and C. Roos. *Flexible aircraft reduced-order LPV model generation from a set of large-scale LTI models*. in *American Control Conference (ACC)*. 2011. IEEE.
6. Poussot-Vassal, C. and F. Demourant, *Dynamical medium (large)-scale model reduction and interpolation with application to aircraft Systems*. AerospaceLab, 2012(4): p. p. 1-11.
7. Theis, J., et al. *Modal Matching for LPV Model Reduction of Aeroservoelastic Vehicles*. in *AIAA Atmospheric Flight Mechanics Conference*. 2015. Kissimmee, Florida.
8. Wang, Y., et al. *Model Order Reduction of Aeroservoelastic Model of Flexible Aircraft*. in *57th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference, AIAA SciTech*. 2016.
9. Zhu, J., et al., *Genetic Algorithm-Based Model Order Reduction of Aeroservoelastic Systems with Consistent States*. Journal of Aircraft, 2017.
10. Theis, J., P. Seiler, and H. Werner. *Model order reduction by parameter-varying oblique projection*. in *American Control Conference (ACC), 2016*. 2016. IEEE.
11. Poussot-Vassal, C. and C. Roos, *Generation of a reduced-order LPV/LFT model from a set of large-scale MIMO LTI flexible aircraft models*. Control Engineering Practice, 2012. **20**(9): p. 919-930.
12. Poussot-Vassal, C. *An iterative SVD-tangential interpolation method for medium-scale MIMO systems approximation with application on flexible aircraft*. in *Decision and Control and European Control Conference (CDC-ECC), 2011 50th IEEE Conference on*. 2011. IEEE.
13. Swei, S.S.-M., G.G. Zhu, and N.T. Nguyen. *Integrated model reduction and control of aircraft with flexible wings*. in *AIAA Guidance, Navigation, and Control (GNC) Conference*. 2013.
14. Al-jiboory, A., et al. *LPV Model Development for a Flexible Wing Aircraft*. in *58th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. 2017.
15. Panzer, H., et al., *Parametric model order reduction by matrix interpolation*. at-Automatisierungstechnik Methoden und Anwendungen der Steuerungs-, Regelungs-und Informationstechnik, 2010. **58**(8): p. 475-484.
16. Grepl, M.A. and A.T. Patera, *A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations*. ESAIM: Mathematical Modelling and Numerical Analysis, 2005. **39**(1): p. 157-181.
17. Paul-Dubois-Taine, A. and D. Amsallem, *An adaptive and efficient greedy procedure for the optimal training of parametric reduced-order models*. International Journal for Numerical Methods in Engineering, 2015. **102**(5): p. 1262-1292.

18. Bui-Thanh, T., K. Willcox, and O. Ghattas, *Model reduction for large-scale systems with high-dimensional parametric input space*. SIAM Journal on Scientific Computing, 2008. **30**(6): p. 3270-3288.
19. Hesthaven, J.S., B. Stamm, and S. Zhang, *Efficient greedy algorithms for high-dimensional parameter spaces with applications to empirical interpolation and reduced basis methods*. ESAIM: Mathematical Modelling and Numerical Analysis, 2014. **48**(1): p. 259-283.
20. Lieberman, C., K. Willcox, and O. Ghattas, *Parameter and state model reduction for large-scale statistical inverse problems*. SIAM Journal on Scientific Computing, 2010. **32**(5): p. 2523-2542.
21. Frangos, M., et al., *Surrogate and reduced-order modeling: a comparison of approaches for large-scale statistical inverse problems [Chapter 7]*. 2010.
22. Forrester, A., A. Sobester, and A. Keane, *Engineering design via surrogate modelling: a practical guide*. 2008: John Wiley & Sons.
23. Koziel, S. and L. Leifsson, *Surrogate-based modeling and optimization*. Applications in Engineering, 2013.
24. Falkiewicz, N.J., et al., *Reduced-order aerothermoelastic framework for hypersonic vehicle control simulation*. AIAA journal, 2011. **49**(8): p. 1625-1646.
25. Crowell, A.R. and J.J. McNamara, *Model reduction of computational aerothermodynamics for hypersonic aerothermoelasticity*. AIAA journal, 2012. **50**(1): p. 74-84.
26. Song, H., et al. *Development of Aeroelastic and Aeroservoelastic Reduced Order Models for Active Structural Control*. in *56th AIAA/ASCE/AHS/ASC Structures, Structural Dynamics, and Materials Conference*. 2015.
27. Lindhorst, K., M. Haupt, and P. Horst, *Aeroelastic Analyses of the High-Reynolds-Number-Aerostructural-Dynamics Configuration Using a Nonlinear Surrogate Model Approach*. AIAA Journal, 2015. **53**(9): p. 2784-2796.
28. Gorissen, D., et al., *A surrogate modeling and adaptive sampling toolbox for computer based design*. Journal of Machine Learning Research, 2010. **11**(Jul): p. 2051-2055.
29. Glaz, B., L. Liu, and P.P. Friedmann, *Reduced-Order Nonlinear Unsteady Aerodynamic Modeling Using a Surrogate-Based Recurrence Framework*. AIAA journal, 2010. **48**(10): p. 2418.
30. Gano, S.E., H. Kim, and D.E. Brown. *Comparison of three surrogate modeling techniques: Datascape, kriging, and second order regression*. in *Proceedings of the 11th AIAA/ISSMO Multidisciplinary Analysis and Optimization Conference, AIAA-2006-7048, Portsmouth, Virginia*. 2006.
31. Lophaven, S.N., H.B. Nielsen, and J. Sondergaard, *DACE: A MATLAB Kriging Toolbox Version 2.0*, in *Technical Univ. of Denmark, Informatics Math and Modelling TR 2002-12*. 2002: Lyngby, Denmark.
32. Meyer, C., et al. *Ground test for vibration control demonstrator*. in *Journal of Physics: Conference Series*. 2016. IOP Publishing.
33. <http://www.mscsoftware.com/product/msc-nastran>.
34. <http://www.zonatech.com/ZAERO.htm>.
35. Pak, C.-g. and S. Truong. *Creating a Test Validated Structural Dynamic Finite Element Model of the X-56A Aircraft*. in *5TH AIAA/ISSMO MULTIDISCIPLINARY ANALYSIS AND OPTIMIZATION CONFERENCE*. 2014.